脑电的重测信度：在多项静息态和任务态实验中的对比*

覃慧怡[1]　丁丽洪[1]　段威[1,2]　雷旭[1]

（[1]西南大学心理学部，重庆 400715）（[2]中国科学院心理研究所，北京 100101）

摘　要　基于脑电探讨心理过程的认知神经机制，需要信号本身具有良好的重测信度。本研究多角度对比了睁眼、闭眼两种静息态脑电以及精神运动警觉任务、Oddball 两种任务态事件相关电位的重测信度。发现静息态的重测信度普遍优于任务态，其中闭眼静息态又比睁眼静息态的重测信度更高，alpha 频段的重测信度在所有频段中最高。对于两种任务态，在时域上均表现为刺激出现后 200 ms 左右较高的重测信度。在空间上，五种状态的结果都表明中央区域比外周区域的重测信度更高，这可能与外周区域更容易出现伪迹有关。本研究涉及多个静息态和任务态脑电实验，在频域、时域、空域三个维度的特征上，较为全面地对比了重测信度并分析了可能的原因，为后续研究脑电信号重测信度选择合适的实验范式和指标提出了建议，对于基础和临床领域的脑电应用都具有重要参考价值。

关键词　脑电，事件相关电位，重测信度，静息态

1 引言

神经影像技术的发展和应用是脑科学研究的重要基础，包括功能磁共振成像（functional magnetic resonance imaging, fMRI）、脑电图（electroencephalography, EEG）和脑磁图（magnetoencephalogram, MEG）等在内的多项脑成像技术受到了心理学、认知神经科学、生物医学工程等领域的广泛关注。相比于 fMRI 和 MEG 等脑成像技术，EEG 因其时间分辨率高、操作简单、无创、便携和价格低廉等多种优势，成为了认知神经科学等领域最常用的研究工具。基于头表脑电信号探讨心理过程的认知神经机制，需要脑电信号具有良好的重测

收稿日期: 2022-09-06

* 国家重点研发计划重点专项(2021YFC2501502)、国家自然科学基金面上项目(NSFC31971028)资助

通信作者: 雷旭, E-mail: xlei@swu.edu.cn

信度(Elliott et al., 2020)。

静息态（通常为睁/闭眼静息态脑电，即 resting-state EEG, rsEEG）和任务态脑电（通常为事件相关电位，即 event related potential, ERP）是头表脑电研究中最常用的两种形式。rsEEG 通常是指人在清醒休息的状态下，在头表记录到的有节律的、反映大脑内在连接的自发神经电活动，它能提供个体有关认知和人格特征差异的相关信息(Deco, Jirsa, & McIntosh, 2011)。因此，许多研究将 rsEEG 扩展到临床研究领域作为识别神经精神疾病的生物标志物(Newson & Thiagarajan, 2019)，例如阿尔茨海默病(Bonanni et al., 2008)、癫痫(Rotondi et al., 2016)、精神分裂症(Siebenhühner et al., 2013)、失眠(Zhao et al., 2021)等。与静息态脑电不同，任务态脑电一般为记录一段认知任务过程中的脑电，也可以是由多个试次叠加后的事件相关电位，即 ERP(Wang et al., 2022)。它反映了人脑在认知过程中的神经电生理的变化，用于探讨个体的认知功能的特点及动态过程，并被应用在脑机接口上(Lugo et al., 2020)。EEG 分析技术包括波形分析、频谱分析、时频分析和源定位等(赵文瑞 等, 2020)。频谱分析采用的频段划分标准通常为：delta（0.5~4 Hz），theta（4~8 Hz），alpha（8~13 Hz），beta (13~30 Hz)和 gamma (30~80 Hz)(Tatum et al., 2016)。本研究重点关注常用的分析技术中的频谱信息和波形特征，前者提供了 rsEEG 在特定频段的功率谱信息，后者关注 ERP 在特定时段脑电成分的振幅和潜伏期等信息。

EEG 在不同场景和时间间隔下的稳定性和一致性一直受到关注。理论计算表明，信度决定了最大可检测到的效度大小，而在心理研究领域中，信度、样本量和效应量之间的相互作用经常被低估(Zuo et al., 2019)。重测信度(test-retest reliability)的研究显得尤为重要。用于评估重测信度的常用统计量是组内相关系数（intra-class correlation coefficients, ICC），通常被定义为个体的方差与总方差的比率。一般情况下，ICC 值小于 0.5 表示信度差，介于 0.5 和 0.75 之间的值表示中等信度，介于 0.75 和 0.9 之间的值表示信度良好，大于 0.9 的值表示信度优良(Koo & Li, 2016)。

近年来多个研究在不同场景中对比了 rsEEG 和 ERP 的重测信度，发现两者各具优势。与 ERP 相比，rsEEG 实验设计简单，易于在特殊疾患群体、特定认知状态中实施。而 ERP 可以设计更有针对性的行为任务，提供更准确的认知过程电生理指标。然而，目前对于 EEG 重测信度的研究大多关注其中一个类别。比如，一些研究者主要关注静息态，探讨在不同的预处理条件、数据长度下的重测信度，发现闭眼条件下的重测信度普遍高于睁眼条件件(Duan et al., 2021)，特定频段如 alpha 和 theta 具有更好的重测信度(Corsi-Cabrera et al., 2007; McEvoy et al., 2000)。另一些研究则关注任务态的重测信度，发现在听觉 Oddball 任务中 ERP 重测相

关性高于 0.9，在 12~16 周的时间间隔内相关性稍低，但是也达到了 0.8(Salinsky et al., 1991)。在精神运动警觉任务（psychomotor vigilance task, PVT）中，发现 ERP 无论是间隔几小时还是几天，重测信度都很好（r 大于 0.9）(McEvoy et al., 2000)。总体来说，已有研究在重测指标、实验设计、数据处理、间隔时长、被试群体等方面有很大的差异，这降低了不同实验范式之间的可比性。例如，尽管以往大多数研究已经发现 rsEEG 中的闭眼状态具有更高的重测信度，但是并未将其与同样测试条件下的 ERP 进行对比。在寻找 ERP 中的最优重测指标时，大多数研究也着力于横向寻找某一个或几个任务中的最优成分，缺少对这几个任务的纵向比较以获得更具普遍参考价值的意见。在空间指标上，对于电极位置的重测信度的研究则更少。对于静息态 MEG 的研究发现，顶叶区域显示出最高的信度(Lew et al., 2021)，但尚不确定将其推广到头表以及 EEG 中的情况。

尽管目前已有多项研究检验了 EEG 的可靠性，但 EEG 在不同实验条件下的可靠性仍缺乏系统的横向比较(Bennett & Miller, 2010; Elliott et al., 2020)。亟待通过实验设计，在同一批被试上对不同任务、不同分析方法进行直接对比。同时，可靠性也取决于预处理步骤和脑功能指标选择，通过横向比较，有利于找到最适合探讨个体差异的指标(Zuo et al., 2019)。在本研究中，我们从更全面的角度讨论了静息态、任务态 EEG 在频域、时域、空域三个维度上的重测信度。我们记录了 42 名健康成年被试分别在间隔 1.5 小时、一个月的三次测试中的 rsEEG 和 ERP，并对两类 EEG 的重测信度进行全方位的比较，试图找出几种实验条件下的总体重测信度分布、最优频段、最优时间点以及最优头表区域。基于先前的研究，我们假设：（1）rsEEG 可能比 ERP 总体重测信度更好；（2）EEG 的信噪比影响重测信度，信噪比越高重测信度越好，因此 rsEEG 的 alpha 频段，ERP 的任务相关成分（时段、脑区），以及受眼电等伪迹影响少的顶叶区域具有更高的重测信度。本研究试图在不同维度的特征上对比重测信度并提出可能的原因，从而为后续脑电实验在选择合适的实验范式和指标提出建议。

2 方法

2.1 被试和实验设计

2.1.1 被试

本实验招募健康被试 42 人，其中男生 14 人，均为右利手，年龄范围 18~26 岁，平均年龄 19.5±1.4 岁。视力正常或矫正正常，无神经系统疾病，头部无创伤。无睡眠障碍，无心理障碍及精神病史，最近一周睡眠质量较好。在脑电实验当天，禁止食用含酒精、含咖啡因的食物或饮料。每个被试在试验后都将得到补偿，在主试向被试详细解释研究方案后，所有

被试均签署了知情同意书。这项研究得到了作者所在单位伦理委员会的批准，所有实验都符合赫尔辛基宣言。

2.1.2 实验设计

每个被试被邀请到实验室两次共进行三次脑电实验。被试在第一次来实验室期间接受了两次脑电实验测试，两次测试间隔 90 分钟。一个月后，被试第二次来实验室，接受第三次测试。在每次脑电实验过程中，所有被试均完成了四种状态，包括两种静息态（睁眼和闭眼）和两种任务态（PVT 和 Oddball 任务）。在每次测试期间，被试首先完成静息态，然后是任务态，两个任务在所有被试中以随机顺序呈现。具体来说，在静息态脑电记录过程中，被试被要求观看屏幕中间的一个白色十字注视点 5 min (EO)；然后闭眼休息 5 min (EC)，他们需要尽可能地保持静止、安静和放松。在 PVT 任务期间，被试需要对屏幕中间出现的计时器以尽可能快的速度做出按键反应，ERP 分析以计时器出现时间为基点（0 时刻）截取脑电信号。在 Oddball 任务期间，被试需要对高频或低频声音进行计数，声音被随机选定为高概率(75%)的标准刺激和低概率(25%)的偏差刺激，ERP 分析以声音刺激出现的时间为基点截取脑电信号。

2.2 EEG 信号采集

EEG 由安装在弹性帽内的 63 个 Ag/AgCl 电极记录，基于扩展的国际 10-20 系统（Brain Products Gmbh，Steingrabenstr，Germany）。其中两个电极用于记录眼电，FCz 用作在线参考电极。采样频率为 500 Hz，采用脑电膏调整电极和头皮间的电阻，所有电极的电阻保持在 5 kΩ 以下。最后，5 种条件被纳入分析，分别是 2 个静息态：睁眼和闭眼，3 个任务态 ERPs：Oddball 的标准条件，偏差条件，以及 PVT。

2.3 rsEEG 预处理与频谱分析

原始 rsEEG 数据的预处理包含 4 个步骤，使用 EEGLAB（版本 2019_1，http://sccn.ucsd.edu）实现。第一步，将原始脑电数据在 0.3~45 Hz 频段范围内进行带通滤波。第二步，人工判读脑电信号识别坏道，然后使用线性插值法来重建被去除的导联信号（替换为周围最近邻通道的平均）。对于所有被试每次脑电实验的数据，坏道的平均数目为 1.61（±1.96， 范围 0~14）个。第三步，对数据进行全脑平均重参考，并将其分割成 4 秒为单位的段（epoch），人工识别筛选并去除坏段。最后保留的分段数量，对于闭眼为：72.56±4.30 段，睁眼为：71.44 ±5.47 段。第四步，进行独立成分分析（ICA）去除眼动对应的伪迹成分。

对于预处理后的 rsEEG 数据集，我们使用 Welch 方法进一步计算了每个电极的频谱。

每个电极的绝对功率进行对数转换以计算功率谱（1 dB = 10×log μV²）。

2.4 ERP 预处理与分析

对于 ERP 数据的预处理步骤和软件同 rsEEG。ERP 分段的事件长度为 1500 ms，包括刺激前 500 ms 和刺激后 1000 ms。对于所有被试的数据，坏道的数目为为 1.82（±1.73，范围 0~9）个。后期分析中，–200~0 ms 用来进行基线校正，一共提取了三种 ERP 脑电，即 PVT、Oddball 偏差刺激和 Oddball 标准刺激。最后保留的分段数量，对于 3 个条件分别为：36.13±8.71 段；12.83±4.40 段和 51.13±12.62 段。三个条件分别进行叠加平均得到 ERP 脑电。

2.5 数据分析

我们使用了类内相关系数 (ICC) 来评估五种实验条件下信号的重测信度。ICC 将被试间变异性与总变异性进行比较，以评估个体重复测量的可靠性(Koo & Li, 2016)。较高的 ICC 值表明被试内部的变异性较低，并且大部分变异性是由被试之间的差异引起的。经过频谱分析，我们计算了五种条件（静息态两种，任务态三种）的 ICC 值。ICC 计算方法基于：

$$MS_p = \frac{d}{n-1}\sum_{i=1}^{n}(\overline{Y_i} - \overline{Y})^2$$

$$MS_e = \frac{1}{(n-1)(d-1)}\sum_{i=1}^{n}\sum_{j=1}^{d}(Y_{ij} - \overline{Y_i} - \overline{Y_j} + \overline{Y})^2$$

本研究中，d（=3）是测试的数量，n（=42）是被试的数量，MS_p 是被试间的均方差，MS_e 是误差的均方差。Y_{ij} 表示第 i 个被试的第 j 次测量值，$\overline{Y_i}$ 表示第 i 个被试所有测量的均值，$\overline{Y_j}$ 表示第 j 次测量时所有被试的均值，$\overline{Y}$ 表示所有测量中所有被试的均值。

$$ICC = \frac{MS_p - MS_e}{MS_p + (d-1)MS_e}$$

我们使用方差分析 (ANOVA) 验证了电极位置和实验任务的交互作用。p 值小于 0.05 被认为具有统计学意义。所有分析中使用 Shapiro-Wilk 检验，Greenhouse-Geisser 校正用于调整任何违反球形检验的影响。

3 结果

3.1 功率谱和 ERP 波形分析

我们首先选取 Cz 电极的信号进行平均，并计算所有被试的平均值和标准差。图 1A 分别显示了在三次实验测量中，睁闭眼 rsEEG 在 0.3~45 Hz 范围内的频谱活动情况。可见闭眼

状态下的绝对功率普遍高于睁眼状态，两种状态的 rsEEG 的最大功率都集中在低频范围内（0.3~13 Hz 左右），并在 10 Hz 左右出现峰值。这对应于 alpha 频段，符合前人的研究结果。图 1B 分别显示了在三种 ERP 中的幅度情况。可见 PVT 任务相比 Oddball 任务诱发了更大的 ERP 成分。PVT 任务在 100 ms 左右诱发了第一个正向波，在 300~700 ms 左右具有最大的幅值。Oddball 任务的两类刺激在 250 ms 前基本重合，均在 130 ms 左右诱发了 N1 波，但是偏差刺激诱发的幅值略大；在 250 ms 左右诱发了 P2 波，偏差刺激的幅值也略大。在 250 ms 之后，两类刺激的 ERP 波形出现较大差异，偏差刺激比标准刺激具有更大的幅度。在 300~400 ms 左右，偏差刺激比标准刺激多出现一个显著的 P3 波。对比三次测试结果可发现，rsEEG 频谱图和 ERP 波形图相似度均较高，且前两次重合度更高。这意味着这 5 种信号的 EEG 在三次实验中都具有良好的重测信度，且短期重测信度优于长期重测信度。



图 1　两种静息态 EEG 和三种任务态 ERP 在 3 次实验（session）中的频谱图（A）和波形图（B）。电极位置均为 Cz，实线表示均值，浅色阴影表示标准误。（A）静息态两种条件的频谱图，横轴代表频率，纵轴代表绝对功率值，紫色线条代表睁眼静息态（EO），绿色线条代表闭眼静息态（EC）。（B）任务态三种条件的波形图，横轴代表时间，纵轴代表幅值，黑色线条代表 PVT 范式，黄色线条代表 Oddball 范式中的偏差刺激（Oddball-D），蓝色线条代表 Oddball 范式中的标准刺激（Oddball-S）。

注：EO: eyes-open; EC: eyes-close; PVT: psychomotor vigilance task; Oddball-D: deviant stimulus of oddball task; Oddball-S: Standard

stimulus of oddball task.

3.2 总体重测信度

对数据质量进行初步检查后，对两种 rsEEG 和三种 ERP 计算类内相关系数 ICC。结果发现（如图 2），rsEEG 的重测信度整体优于 ERP。其中，两种静息态的 ICC 值都集中在 0.5 周围。EO 状态下主要在 0.45 和 0.6 有聚集，EC 状态下在 0.4 和 0.6 有聚集，两者的 ICC 范围大体相似（图 2A），EC 状态下的 ICC 值在均值、最大值、中值三个指标上均优于 EO 状态（表 1），配对样本 t 检验差异显著（$t=-2.71, p=0.007$）。而在 ERP 中，PVT 的重测信度在均值、最大值、中值三个指标上均优于 Oddball 范式（表 1）。PVT 的 ICC 值集中在 0.5 左右且整体分布更集中，Oddball 下的 ICC 值集中在 0.35 左右（图 2B）。对三种 ERP 信号的 ICC 值进行单因素方差分析，结果显示差异显著（$F=753.16, p<0.001$），事后检验结果显示 PVT 条件的 ICC 值与 Oddball 两种条件的 ICC 值差异显著，Oddball 两种条件的 ICC 值差异不显著。

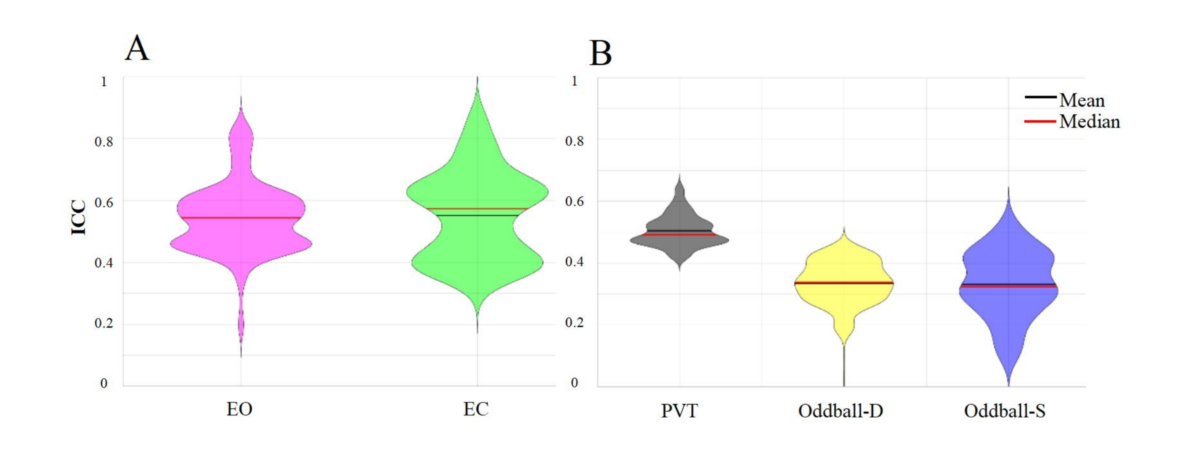


图 2　ICC 值在 5 种条件的功率（A）和幅值（B）中的分布情况，包括所有电极、频率点（rsEEG）和时间点(ERPs)。横轴代表条件，纵轴代表 ICC 值。黑色参考线代表平均值，红色参考线代表中值。（A）静息态两种条件下 ICC 值的分布情况。（B）任务态三种条件下 ICC 值的分布情况。

注：EO: eyes-open; EC:eyes-close; PVT: psychomotor vigilance task; Oddball-D: deviant stimulus of oddball task; Oddball-S: Standard stimulus of oddball task.

表 1　5 种条件下 ICC 的整体分布情况

	EO	EC	PVT	Oddball-D	Oddball-S
均值±标准差	0.54±0.16	0.55±0.19	0.50±0.13	0.33±0.13	0.33±0.16

最大值	0.91	0.94	0.82	0.68	0.80
中值	0.53	0.56	0.51	0.35	0.35
差异检验	EC > EO, $t = -2.71$, $p = 0.007$		PVT > Oddball-D / Oddball-S, $F = 753.16$, $p < 0.001$		

注：差异检验对静息态和任务态分别进行统计比较，对静息态两种条件下的 ICC 均值进行配对样本 t 检验，对任务态三种条件下的 ICC 均值进行单因素被试内 ANOVA 分析，“>”表示有显著差异（$p < 0.05$），“/”表示无显著差异（$p >= 0.05$）.

3.3 时频域下的重测信度

在静息态频域范围内和任务态时域范围内所有电极的 ICC 进行平均，以分别评估 rsEEG 和 ERPs 的重测信度在频、时域上的变换趋势。结果发现，在静息态条件下两种 rsEEG 的最大 ICC 值都出现在 alpha（8~12 Hz）频段（ICC > 0.8），其他频段的 ICC 值集中在 0.6 和 0.4 左右。同时值得注意的是，在 0 ~ 23 Hz 频率范围内 EC 的 ICC 值更大，在 23 ~ 45 Hz 频率范围内 EO 的 ICC 值更大（图 3A）。在任务态条件下，两种实验范式下的最大 ICC 值都出现在 200 ms 左右（200±20 ms）($ICC_{PVT} > 0.6$, $ICC_{Oddball-S} > 0.5$, $ICC_{Oddball-D} > 0.4$)，其他时间点的 ICC 值略有降低。考虑到在计算 ERP 成分时，通常按顶点的幅值来进行分析，因而我们又对 PVT 中的 N1 成分（130~200ms）(Hoedlmoser et al., 2011)和 Oddball 任务中的 P200 成分（150~275ms）(赵文瑞 等, 2020)分别计算了 ICC 值。从结果可见，对成分的 ICC 计算结果与本研究中 200ms 左右的结果基本一致。PVT、Oddball-D 和 Oddball-S 分别得到的 ICC 为：0.67、0.40 和 0.28。同时值得注意的是，在所有时间范围内，PVT 范式的 ICC 值均大于 Oddball 范式。在 Oddball 范式的两类刺激中，0 ~ 400 ms 范围内标准刺激诱发的 ICC 更大，400 ms 之后偏差刺激诱发的 ICC 更大（图 3B）。

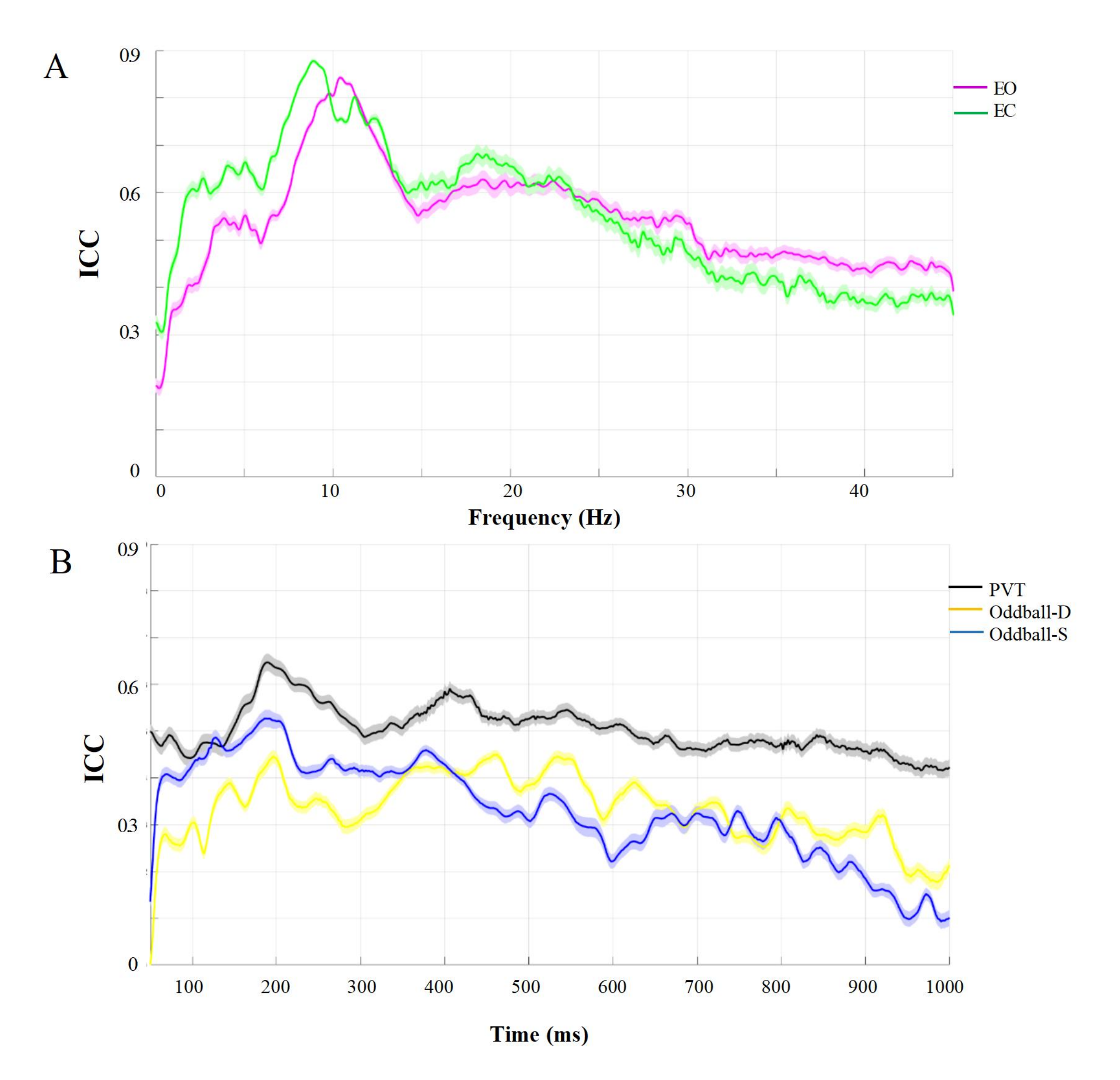


图 3　**所有电极** ICC 值在 5 种条件中的变化情况。实线表示均值，浅色阴影表示标准误。（A）静息态两种条件下 ICC 值的变化，横轴代表频率，纵轴代表 ICC 值，紫色线条代表睁眼静息态（EO），绿色线条代表闭眼静息态（EC）。（B）任务态三种条件下 ICC 值的变化，横轴代表时间，纵轴代表 ICC 值，黑色线条代表 PVT 范式，黄色线条代表 Oddball 范式中的偏差刺激（Oddball-D），蓝色线条代表 Oddball 范式中的标准刺激（Oddball-S）。

注：EO: eyes-open; EC: eyes-close; PVT: psychomotor vigilance task; Oddball-D: deviant stimulus of oddball task; Oddball-S: Standard stimulus of oddball task.

3.4 重测信度的空间分布

从上面的分析中可以看到，静息态最大的 ICC 值出现在 alpha 频段（8 ~ 12 Hz），任务态最大的 ICC 值出现在 200 ms 左右（±20 ms）。在进行空间位置的对比时，为了减小其他因素的影响，更直观地对比不同实验状态、条件下的 ICC 值，我们分别选取了 5 种信号在时频域中最大 ICC 值出现的范围（alpha 频段、200 ms 左右）。如图 4 所示，对于静息态，

EO 条件下较高的 ICC 值集中出现在中央区和顶叶，EC 条件下的 ICC 值总体更大、分布也更广。同时，静息态的两种条件对应的高 ICC 值区域都集中在中央区（图 4）。对于任务态，PVT 的高 ICC 值区域出现在额叶、中央区和顶叶，Oddball 标准刺激的 ICC 值更大，两种刺激下的高 ICC 值区类似，都集中在额叶、中央区和颞叶（图 4）。

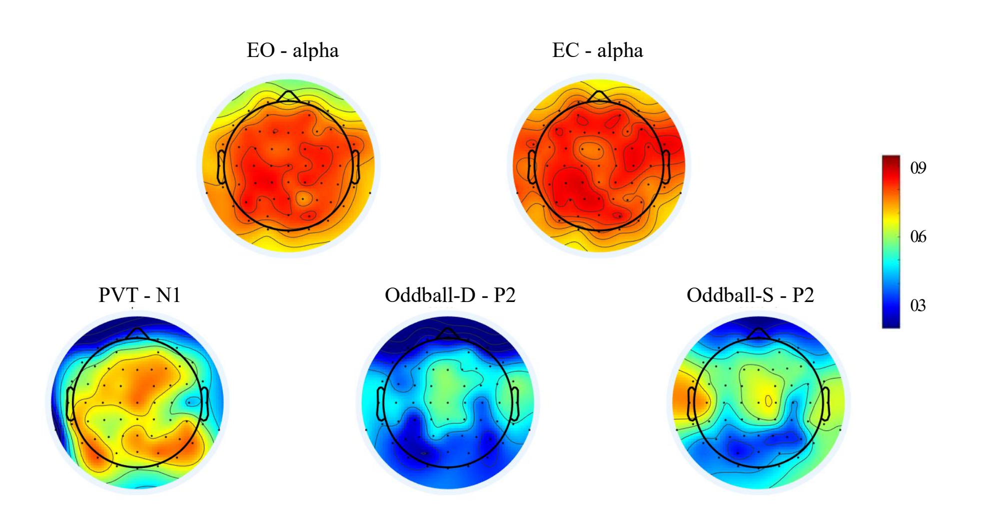


图 4　rsEEG 的 alpha 频段和 ERP 的 P2 成分对应的 ICC 值在头表的分布.

注：EO: eyes-open; EC: eyes-close; PVT: psychomotor vigilance task; Oddball-D: deviant stimulus of oddball task; Oddball-S: Standard stimulus of oddball task.

对比五种条件最大 ICC 值头表拓扑分布，可以发现前三种条件下的最大 ICC 值均主要集中出现在中央区域。而在 Oddball 除了中央区域对应更大的 ICC 值外，颞叶的 ICC 值也高于其他区域。基于信噪比的理论，我们假设前三种条件之所以中央区域 ICC 更大，是因为这一区域受到更少的噪声伪迹（眼电、心电、肌电等）影响，从而具有更高的信噪比。而 Oddball 条件下，颞叶有更多激活，使颞叶区域的电极采集到的 EEG 幅度更大，因而信噪比更高、ICC 更大。为了验证这个假设，我们将头表电极划分为受噪声影响少的中央区域和影响多的外周区域。如图 5，外周区域包括（AFz, FPz, FP1 等）共 30 个电极，其余为中央区域。分别计算了两个区域的平均 ICC 值（图 5），将电极位置与实验条件进行两因素 ANOVA 分析，再进行独立样本 T 检验。ANOVA 分析结果显示，区域位置因素的主效应显著（$F(1, 23)=18.58$，$p<0.001$），中央比外周的 ICC 高 0.068（$p<0.001$）。条件因素对 ICC 的影响也具有统计学意义（$F(4, 92)=260.01$，$p<0.001$）。因为条件因素有 5 个水平，所以对 5 个水平条件下的 ICC 进行两两比较，发现所有条件下 ICC 的差异均具有统计学意义($p<0.001$)。ANOVA 分析的结果证明了电极区域位置和实验条件主效应显著，但交互作用对 ICC 值无显

著的影响。之后进行事后检验，独立样本 T 检验结果显示，EO、EC、PVT、Oddball-D 四种条件下中央区域和外周区域电极的 ICC 值差异显著，但 Oddball-S 条件下差异不显著（表 2）。

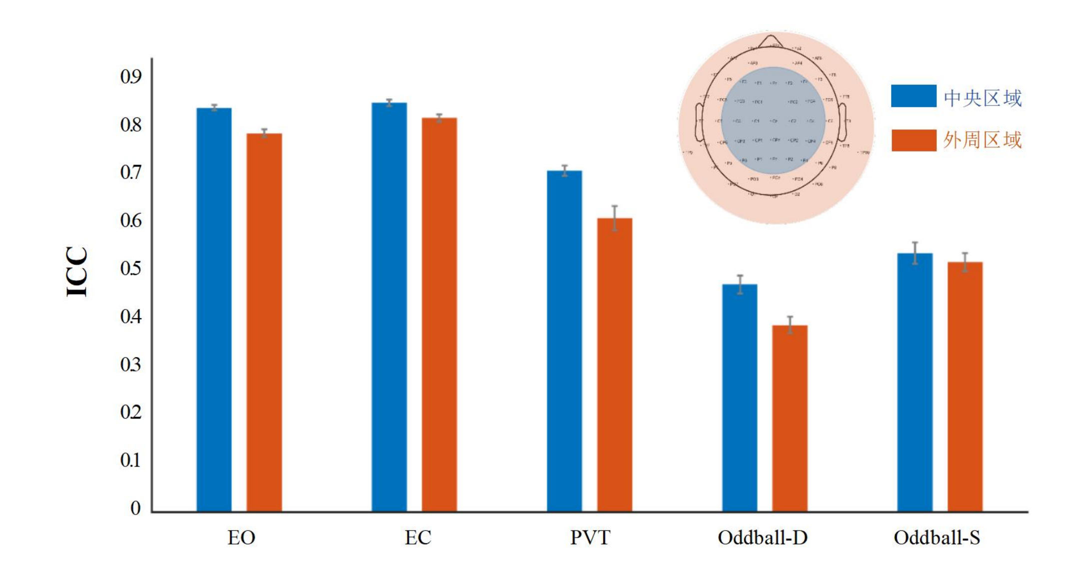


图 5　5 种条件（EO，EC，PVT，Oddball-D，Oddball-S）下 2 个感兴趣区（中央和外周区域）的 ICC 平均值。横轴代表条件，纵轴代表 ICC 值。蓝条代表中央区域，橙条代表外周区域。

注：插入图为中央与外周区域电极划分情况, 蓝色为中央区域，橙色为外周区域.

表 2　5 种条件下 2 个感兴趣区的 ICC 值及其独立样本 T 检验结果

实验条件	EO		EC		PVT		Oddball-D		Oddball-S	
感兴趣区	中央	外周	中央	外周	中央	外周	中央	外周	中央	外周
ICC 平均值	0.81	0.76	0.82	0.79	0.69	0.59	0.46	0.38	0.52	0.50
ICC 方差	0.0007	0.0021	0.0010	0.0017	0.0026	0.0222	0.0072	0.0100	0.0106	0.0127
t	5.42		3.10		3.52		3.26		0.44	
p	＜0.001		0.003		0.001		0.002		0.661	

注：EO: eyes-open; EC: eyes-close; PVT: psychomotor vigilance task; Oddball-D: deviant stimulus of oddball task; Oddball-S: Standard stimulus of oddball task.

4 讨论

通过采用多种任务态和静息态实验，本研究对比了五种条件的 EEG 数据在时域、频域、

空域的重测信度。我们发现 rsEEG 比 ERP 的 ICC 值普遍更高。静息态闭眼状态的 ICC 比睁眼状态更高，这与前人的研究结果一致(Corsi-Cabrera et al., 2007; Duan et al., 2021)。无论是在睁眼还是闭眼状态下 alpha 频段的重测信度均优于其他频段。任务态条件下 PVT 的 ICC 比 Oddball 都高，并且两种实验范式诱发的最大 ICC 值均在 200 ms 左右。在空间位置上除了 Oddball 标准刺激条件，其他条件下的最大 ICC 值都分布在中央区域。据我们所知，本研究是首次从时域、频域、空域多维度对比脑电重测信度的研究，并发现多个有利于提高重测信度的条件。

4.1 静息态功率谱的重测性

从 rsEEG 的频谱图中可以看出，对于两种 rsEEG 而言 alpha 频段的功率值都最大，而 alpha 的活动主要反映了对任务无关神经活动的抑制(Klimesch et al., 2007; Uusberg et al., 2013)，这与以 alpha 频段为主的闭眼静息态具有的高重测信度的结果一致。

对比两种 rsEEG，发现其 ICC 分布范围较大且比较相似，最大值均大于 0.9。对于 30 天定量 EEG 的研究也发现静息状态下有良好的重测信度，其中闭眼状态的 ICC 平均值达到了 0.81(Cannon et al., 2012)。在我们的结果中，闭眼条件相比于睁眼条件重测信度更高，这一结果与之前的研究一致(Corsi-Cabrera et al., 2007)。我们根据影响数据质量的信噪比理论分析(Zuo et al., 2019)，除了这两种条件间 EEG 活动的固有差异外，还有两个重要因素影响其重测信度。首先，从信号的角度，闭眼条件下 alpha 频段的活动比睁眼条件更大，在功率谱上信号能量更大。同时，alpha 节律与来自全脑的认知活动有关，可以在广泛区域的电极中被检测到，而其他高频活动（如 beta 和 gamma 节律）更多地与大脑局部活动有关(Martín - Buro et al., 2016)，其在全脑的影响被削弱了。其次，从噪声的角度这两种情况受眼动伪迹影响的程度不一样，闭眼状态下眼动伪迹明显少于睁眼状态，噪声的能量更小(Ding et al., 2022)。

4.2 任务态 ERP 成分的重测性

对 ERP 成分的分析发现，在 PVT 范式中 P1(在刺激开始后约 100 ms 达到峰值的正向成分)和 N1(在刺激开始后约 160 ms 达到谷峰值的负向成分)是与选择性、反射性、视觉空间注意以及特征检测密切相关的成分(Hoedlmoser et al., 2011)。在 Oddball 范式中，标准刺激诱发 N1、P2 波，而偏差刺激诱发 N1、P2、N2、P3 和 N4 波（图 1）。我们的 ERP 波形图与先前的研究基本一致(Cassidy et al., 2012)，初步证实了数据的可靠性。

基于以往对于脑电重测信度的研究，我们最初假设幅度值较大的成分在重测信度上表现更好，然而结果并不完全支持该假设。对于三种 ERP 信号而言，重测信度最好的成分都出

现在 200 ms 左右。从图 1 的波形图中可以看出，三种信号都在 200 ms 左右出现了任务相关成分（PVT 范式诱发的 N1 成分，Oddball 范式诱发的 P2 成分），这部分结果符合基于信噪比理论提出的：有效信息占比更大重测信度更高的假设(Zuo et al., 2019)。先前对于多种任务的研究也发现，潜伏期较短的成分（P1、N1、N170、ERN）的峰值振幅最稳定(Cassidy et al., 2012)。所以基于脑电采集方式和信噪比理论分析，在 ERP 早期成分中来自全脑的信号反应了更加同步的少数心理认知过程。这种信号同步在之后多个心理认知加入的过程中被削弱，因此 ERP 中越晚期的成分所反应的全脑神经活动越复杂，在重测时可能就难以保持一致性。

4.3 空间特征

在寻找最优重测性的电极位置时，目前主要有两种方法：第一种是选择受眼电、肌电等伪迹影响较小的电极位置，第二种是选择与任务更相关的大脑区域对应的电极位置。这两种方法分别从减少噪音和增加信号两个方面提高了数据的信噪比，从而提高了重测信度。我们基于“信噪比”理论提出假设，旨在对比信息和噪声两者对于重测信度影响的差异。从减少噪声的角度而言，因为中央区域采集的 EEG 相比于外周区域更少受到来自眼动、头动、心跳以及电极连接不牢固的影响，所以信噪比更高。

我们发现所有实验条件下最大 ICC 值都集中地分布在中央区域，且中央电极比外周电极的重测信度都高（除了 Oddball 标准刺激条件），初步证实了目前实验中噪声伪迹对于重测信度的影响较大。Oddball 标准刺激条件是个例外，从图 1B 的波形图可以看到，该条件仅在 200 ms 左右诱发了一个显著波峰，而 P3a（220～280 ms）成分反映了听觉刺激驱动的自下而上的前脑区注意加工机制，激活脑区位于初级听觉皮质和前额叶(Vanhaudenhuyse et al., 2008)，意味着 Oddball 标准刺激条件的 ERP 信号更多地分布在外周区域。因此，除了选择受伪迹干扰小的电极外，考虑实验任务本身诱发的 ERP 成分以提高信号强度也是提高重测信度的有效手段。值得注意的是，随着电极性能的提高和预处理技术的进步，目前发现的周边区域相对中央区域的劣势在未来可能会得到改善。

4.4 多种任务态和静息态的对比

为了得到更普遍适用的指标，本研究选择的实验是在大多数脑电实验中采用的静息态睁眼、闭眼状态，以及两种练习效应较小的任务态范式，以确保在多次重复测量的过程中认知心理活动的一致性。虽然结果发现静息态脑电比任务态脑电的 ICC 值更高（图 2），但结果的推广需要谨慎。有学者提出 rsEEG 的可靠性低于 ERP，原因是静息态下的注意力和警惕性水平比任务执行时的变化更大，而执行任务过程时的精神状态通常更为统一，因此对 EEG

有稳定作用(Burgess & Gruzelier, 1993)。但近些年越来越多的研究发现，rsEEG 的特定频段有比 ERP 更高的重测信度(Shirk et al., 2017)。我们推测这可能基于某些频段代表了某种恒定的认知神经活动(Feyissa & Tatum, 2019)。另外近年来 rsEEG 预处理技术的进步也大大减少了来自伪迹的干扰(Duan et al., 2021)。

值得注意的是，PVT 的 ICC 集中在 0.5 左右，最大值约为 0.8，这个结果与静息态条件接近（图 2）。早期的研究表明 PVT 具有高可靠性，在某些电极和频段中甚至高于静息态(McEvoy et al., 2000)。尽管有研究对 Oddball 范式下的 ERP 进行了测量并得到了较高的重测性(Salinsky et al., 1991)，但是本研究发现 Oddball 的 ICC 结果最差，中值在 0.35 左右。从实践的角度来看，研究人员所能做的就是尽最大努力测量和消除非感兴趣信号（如运动伪迹等）来提高信噪比，并使用更优的方法提高信效度(Zuo et al., 2019)。

此外，在本研究还发现试次数似乎和重测信度相关：闭眼 rsEEG 试次数最多，对应了最大的 ICC，Oddball 偏差条件的试次数最少对应了最小的 ICC。与先前的研究一致(Cassidy et al., 2012; Ding et al., 2022; Duan et al., 2021)，我们推测试次数可能通过影响信噪比间接影响重测信度。在任务态和静息态 EEG 实验中增加试次数也许是提高重测信度的有效方式。

4.5 局限性

本研究广泛对比了多种任务态和静息态的脑电重测特性，但是在某些方面仍存在缺陷。第一，选择了传统的数据处理方法，这虽然使得结果在更多的研究中具有可比性，但考虑到目前对于预处理方法和参数影响脑电信号重测信度的结论尚不一致(Duan et al., 2021; Suarez-Revelo et al., 2016)。本研究的结论对于区别较大的预处理策略和许多新方法（如功能连接，脑网络等）的参考意义有限，未来的研究也应更多着眼于新策略新方法，推动脑电数据处理和分析流程的标准化(Cohen, 2017)。第二，对 EEG 的分析仅局限在头表水平，并没有对其进行源分析，这可能削弱了其认知神经方面的理论支持，源分析角度的重测情况可以参考最新的研究报告(Ding et al., 2022; Duan et al., 2021)。第三，在对比静息态与任务态时，时长和试次数也是需要被考虑的问题，而两种状态中实验时长和坏段比例通常不同，导致两种状态的试次数并不统一。第四，两种状态的难度也有差异，而本研究局限于静息态以及与注意力、警惕性相关的两种任务态，将此结论推广到更广泛的自然刺激实验、任务态实验范式时需谨慎。第五，本实验被试群体仅限于健康青年人，尚需在更加多样化的群体中进行验证。

5 结论

综上所述，脑电重测性在多个实验条件、时频域指标、空间位置上都具有一定的可比性，平均 ICC 重测信度维持在 0.33 到 0.55 之间。静息态比任务态具有更高的重测信度，其中闭眼 rsEEG 的重测信度最好，PVT 任务所反应的精神警觉性也有较高重测稳定性，这些都可以用试次数的增加和稳定的认知状态带来的高信噪比来解释。在频域上，静息态 alpha 频段的重测信度最高；在时域上，任务态 200 ms 左右的成分最稳定；在空间上，中央区域的重测信度普遍优于外周区域，时频空三个维度的证据初步支持了信噪比对重测信度的决定作用。总之，我们的研究结果在更广泛的因素上讨论了重测信度。这些结果对研究人员选择实验范式和指标提出了建议，对于基础和临床领域的脑电应用都具有重要参考价值。

参考文献

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences, 1191*(1), 133–155.

Bonanni, L., Thomas, A., Tiraboschi, P., Perfetti, B., Varanese, S., & Onofrj, M. (2008). EEG comparisons in early Alzheimer's disease, dementia with Lewy bodies and Parkinson's disease with dementia patients with a 2-year follow-up. *Brain, 131*(3), 690–705.

Burgess, A., & Gruzelier, J. (1993). Individual reliability of amplitude distribution in topographical mapping of EEG. *Electroencephalography and Clinical Neurophysiology, 86*(4), 219–223.

Cannon, R. L., Baldwin, D. R., Shaw, T. L., Diloreto, D. J., Phillips, S. M., Scruggs, A. M., & Riehl, T. C. (2012). Reliability of quantitative EEG (qEEG) measures and LORETA current source density at 30 days. *Neuroscience Letters, 518*(1), 27–31.

Cassidy, S. M., Robertson, I. H., & O'Connell, R. G. (2012). Retest reliability of event-related potentials: evidence from a variety of paradigms. *Psychophysiology, 49*(5), 659–664.

Cohen, M. X. (2017). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *International Journal of Psychophysiology, 111*, 80–87.

Corsi-Cabrera, M., Galindo-Vilchis, L., del-Río-Portilla, Y., Arce, C., & Ramos-Loyo, J. (2007). Within-subject reliability and inter-session stability of EEG power and coherent activity in women evaluated monthly over nine months. *Clinical Neurophysiology, 118*(1), 9–21.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience, 12*(1), 43–56.

Ding, L., Duan, W., Wang, Y., & Lei, X. (2022). Test-retest reproducibility comparison in resting and the mental task states: A sensor and source-level EEG spectral analysis. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology, 173*, 20–28.

Duan, W., Chen, X., Wang, Y. J., Zhao, W., Yuan, H., & Lei, X. (2021). Reproducibility of power spectrum, functional connectivity and network construction in resting-state EEG. *Journal of neuroscience methods, 348*, 108985.

Elliott, M., Knodt, A., Ireland, D., Morris, M., Poulton, R., Ramrakha, S., . . . Hariri, A. (2020). What is the test-retest reliability of common task-fMRI measures? New empirical evidence and a meta-analysis. *Biological Psychiatry, 87*(9), S132–S133.

Feyissa, A. M., & Tatum, W. O. (2019). Adult EEG. *Handbook of Clinical Neurology, 160*, 103–124.

Hoedlmoser, K., Griessenberger, H., Fellinger, R., Freunberger, R., Klimesch, W., Gruber, W., & Schabus, M. (2011). Event-related activity and phase locking during a psychomotor vigilance task over the course of sleep deprivation. *Journal of Sleep Research, 20*(3), 377–385.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Reviews, 53*(1), 63–88.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine, 15*(2), 155–163.

Lew, B. J., Fitzgerald, E. E., Ott, L. R., Penhale, S. H., & Wilson, T. W. (2021). Three-year reliability of MEG resting-state oscillatory power. *Neuroimage, 243*, 118516.

Lugo, Z. R., Pokorny, C., Pellas, F., Noirhomme, Q., Laureys, S., Müller-Putz, G., & Kübler, A. (2020). Mental imagery for brain-computer interface control and communication in non-responsive individuals. *Annals of Physical and Rehabilitation Medicine, 63*(1), 21–27.

Martín - Buro, M. C., Garcés, P., & Maestú, F. (2016). Test - retest reliability of resting - state magnetoencephalography power in sensor and source space. *Human Brain Mapping, 37*(1), 179–190.

McEvoy, L. K., Smith, M. E., & Gevins, A. (2000). Test-retest reliability of cognitive EEG. *Clinical Neurophysiology, 111*(3), 457–463.

Newson, J. J., & Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: a review of resting state studies. *Frontiers In Human Neuroscience, 12*, 521.

Rotondi, F., Franceschetti, S., Avanzini, G., & Panzica, F. (2016). Altered EEG resting-state effective connectivity in drug-naïve childhood absence epilepsy. *Clinical Neurophysiology, 127*(2), 1130–1137.

Salinsky, M. C., Oken, B. S., & Morehead, L. (1991). Test-retest reliability in EEG frequency analysis. *Electroencephalography and Clinical Neurophysiology, 79*(5), 382–392.

Shirk, S. D., McLaren, D. G., Bloomfield, J. S., Powers, A., Duffy, A., Mitchell, M. B., . . . Atri, A. (2017). Inter-Rater Reliability of Preprocessing EEG Data: Impact of Subjective Artifact Removal on Associative Memory Task ERP Results. *Frontiers In Neuroscience, 11*, 322.

Siebenhühner, F., Weiss, S. A., Coppola, R., Weinberger, D. R., & Bassett, D. S. (2013). Intra-and inter-frequency brain network structure in health and schizophrenia. *Plos One, 8*(8), e72351.

Suarez-Revelo, J., Ochoa-Gomez, J., & Duque-Grajales, J. (2016). Improving test-retest reliability of quantitative electroencephalography using different preprocessing approaches. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2016*, 961–964.

Tatum, W. O., Olga, S., Ochoa, J. G., Munger Clary, H., Cheek, J., Drislane, F., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 7: Guidelines for EEG Reporting. *Journal Of Clinical Neurophysiology, 33*(4), 328–332.

Uusberg, A., Uibo, H., Kreegipuu, K., & Allik, J. (2013). EEG alpha and cortical inhibition in affective attention. *International Journal of Psychophysiology, 89*(1), 26–36.

Vanhaudenhuyse, A., Laureys, S., & Perrin, F. (2008). Cognitive event-related potentials in comatose and post-comatose states. *Neurocritical Care, 8*(2), 262–270.

Wang, Y., Duan, W., Dong, D., Ding, L., & Lei, X. (2022). A test-retest resting, and cognitive state EEG dataset during multiple subject-driven states. *Scientific data, 9*(1), 566.

Zhao W. R., Li C. Y., Chen J. J., & Lei, X. (2020). Insomnia disorder and hyperarousal: evidence from resting-state and sleeping EEG．*Scientia Sinica (Vitae)*，*50*(3) : 270–286．
[赵文瑞, 李陈渝, 陈军君, 雷旭. (2020). 失眠障碍与过度觉醒:来自静息态脑电和睡眠脑电的证据. *中国科学:生命科学, 50*(03), 270–286.]
Zhao, W., Van Someren, E. J. W., Li, C., Chen, X., Gui, W., Tian, Y., . . . Lei, X. (2021). EEG spectral analysis in insomnia disorder: A systematic review and meta-analysis. *Sleep Medicine Reviews, 59*, 101457.
Zuo, X. N., Xu, T., & Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nature Human Behaviour, 3*(8), 768–771.

Test-retest reliability of EEG: A comparison across multiple resting-state and task-state experiments

QIN Huiyi[1], DING Lihong[1], DUAN Wei[1,2], LEI Xu[1]

([1] Faculty of Psychology, Southwest University, Chongqing 400715, China)

([2] Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)

Abstract

Owing to its advantages in time resolution, electroencephalography (EEG) provides an important basis for studying the dynamic cognitive process of the human brain. To explore the electrophysiological mechanism of psychological processes, scalp EEG must have good test-retest reliability. Most studies explore the reliability of the resting-state EEG (rsEEG) or event-related potentials (ERP), lacking a comprehensive comparison of multiple states. We comprehensively compared the test-retest reliability of the two rsEEG with eyes-open (EO) and eyes-close (EC) states, and the ERPs of PVT and oddball tasks, from frequency, time, and spatial domains to identify more widely applicable indicators.

A total of 42 healthy adults (age range = 18 – 26 years old; mean = 19.5 ± 1.4 years old; 14 males) underwent all three EEG recording sessions, including the present (Session 1), 90 mins later (Session 2), and one month later (Session 3). During each EEG recording session, all the participants completed the same five states including two resting states (eyes-open, eyes-closed, each with 5 minutes) and two task states (PVT and oddball task). Intra-class correlation

coefficients (ICC) were computed to assess the test-retest reproducibility of the five states.

This study compares and analyses the test-retest reliability of two resting-state and three task-state EEG from the perspectives of time, frequency, and spatial domains. Results revealed the following: (1) The test-retest reliability of rsEEG was generally better than that of ERP. (2) For rsEEG, the test-retest reliability of the EC resting-state was higher than that of the EO, with the ICC median value of approximately 0.6. Furthermore, the test-retest reliability of the alpha band was the highest in all frequency bands. (3) For the two task-states ERP, the overall ICC of the PVT paradigm was higher than that of the oddball paradigm, and the test-retest reliability was highest at about 200 ms after the stimulus onsets. (4) In the spatial domain, the test-retest reliability is higher in the central region than in the peripheral region, which may be related to the signal-to-noise ratio (SNR).

Our research involves multiple resting-state and task-state experiments. Based on the characteristics of frequency, time, and space domains, we comprehensively compared the optimal retest characteristics of multiple EEG and suggest the possible reasons. Some suggestions for the selection of appropriate experimental paradigms and indicators for the follow-up study of EEG test-retest reliability are provided and guide the application of EEG in the basic and clinical fields.

Key words EEG, event-related potential, test-retest reliability, resting-state EEG